

A map of the state of Georgia, showing all 159 counties. The counties are labeled with their names. The text "GaNCH Project Partners Meeting" is overlaid on the map in a large, black, sans-serif font. The county of DeKalb is highlighted in orange.

GaNCH Project Partners Meeting

2020-01-30

Agenda

- Introductions and project overview
- Introductions of project partners
- Project update
- Challenges
- Future Steps
- Next Meeting
- Questions



Project Team



- Christine Wiseman
- Co-PI
- Department Head, Digital Services Department



- Cliff Landis
- Co-PI
- Digital Initiatives Librarian



- Allyson F. Smith
- Graduate Student Assistant



- Matthew Stephens
- Web Developer



- Jessica Leming
- Digital Curation Librarian



- Alex Dade
- Library Technical Specialist

Project Overview

- This one year project will create a **publicly editable directory** of Georgia's Natural, Cultural and Historical Organizations (NCHs)
- This will allow for quick retrieval of **location and contact information for disaster response**.
- Directory information will be compiled, updated, and uploaded to **Wikidata**, the linked open data database from the Wikimedia Foundation.
- **A website will be created** that can display embedded subsets of the directory for emergency responders (by city or county).
- The website will include a semi-automated feature to email organizations, in order to provide **annual updates to directory information**.
- The project is being funded by a **LYRASIS Catalyst Grant**.



Our Partner Organizations



Georgia Historical Records
Advisory Council



Atlanta University Center
Robert W. Woodruff Library



Our Partner Contacts

Partner Organization	Contact Person
LYRASIS	Jessica Crouch
Heritage Emergency Response Alliance (HERA)	Tina Seetoo
	Sigourney Smuts
Georgia Public Library Service (GPLS), Georgia HomePLACE	Angela Stanley
Digital Library of Georgia (DLG)	Sheila A McAlister
Georgia Association of Museums and Galleries	Matt Davis
Georgia Historical Records Advisory Council (GHRAC)	Glenda Anderson
Savannah Heritage Emergency Response (SHER)	Laura Waller
Performing Arts Readiness	Steve Eberhardt
South Arts	Mollie Quinlan-Hayes
Georgia Emergency Management and Homeland Security Agency (GEMA/HS)	Venessa Sims
Georgia Department of Natural Resources (DNR), Historic Preservation Division	Bill Hover
Georgia Council for the Arts	Karen Paty
Society of Georgia Archivists - Outreach Committee "Hidden Archives" Project	Katie Twomey

Project Update: Datasets

- 14 source datasets
- **3 ingested (GAMAG, GPSLV, SAGIS)**
- 2 in-process (GHRAC, GHSOC)
- 1 “OTHER” catch-all

Shortcode	Organization	Source Name	Link	# Items
GHRAC	Georgia Historical Records Advisory Council	Directory of Historical and Cultural Organizations	https://apps.georgiaarchives.org/GHRAB/	641
GHSOC	Georgia Historical Society	Affiliate Chapter List by County	https://georgiahistory.com/education-outreach/affiliate-chapter-program/affiliate-chapter-list-by-county/	44+36+47 = 127
GAMAG	Georgia Association of Museums and Galleries	Museum Links	http://www.gamg.org/museumlinks.html	155
GPSLV	Georgia Public Library Service	All Public Library Facilities	https://georgialibraries.org/allpubliclibrary/facilities/	392
GALIL	GALILEO	Participating Institutions	https://about.galileo.usg.edu/institutions	93
TC SOG	Technical College System of Georgia	Find a College	https://tcsg.edu/find-a-college/	22
ARTGA	ArtsGeorgia	Directory of Directories	https://www.artsgorgia.net/directory/directory-of-directories/calendars/	Meta
GSCCC	Georgia Superior Court Clerks' Cooperative Authority	Clerks	https://www.gscca.org/Clerks/	159
GINFO	GeorgiaInfo	Counties	https://georgiainfo.galileo.usg.edu/topics/counties/	159
IMLSS	Institute of Museum and Library Services	Georgia Museums List	https://docs.google.com/spreadsheets/d/104wXi8w6VXl1JFjNYltpHh9G7JTtztVD-ZjhGF70/	491
GADNR	Georgia Department of Natural Resources	Georgia State Parks, Historic Sites, and Partners	https://gastateparks.org/Map	72
GCFTA	Georgia Council for the Arts	FY2020 Grant Applicants	N/A (emailed)	556
NCCSD	Urban Institute	National Center for Charitable Statistics Data Archive	https://nccs-data.urban.org/index.php	???
EOBMF	Internal Revenue Service	Exempt Organizations Business Master File Extract (EO BMF)	https://www.irs.gov/charities-non-profits/exempt-organizations-business-master-file-extract-EO-BMF	???
SAGIS	Savannah Area GIS	SAGIS	https://www.sagis.org/	61
GAPAR	LYRASIS	Georgia Performing Arts Readiness	Emailed	15
OTHER	One-off organizations	Email	N/A	24

https://github.com/clifflandis/GaNCH/blob/master/data/data_sources.md

Project Update: Index of all organizations

- Current [index](#) (de-duped count) is at **1,131** (was at 558 last time we met)!
- Because our “fuzzy lookup” deduplication, we’re finding occasional duplicates that were missed (based on alternative spellings and organizational name changes)

Baldwin Library		0.0000
Banks County Historical Society	0459 Banks County Public Library	0.6759
Barnesville-Lamar County Historical Society	0315 Barnesville-Lamar County Library	0.8800
Barrington Hall	0012 Barrington Hall	1.0000
Barrow County Historical Society and Museum	0083 Lowndes County Historical Society and Museum	0.6861
Bartow County Genealogical Society	0037 Decatur County Historical and Genealogical Society	0.5962
Cartersville Public Library	0474 Yatesville Public Library	0.8000
Bartow History Museum	0013 Bartow History Museum	1.0000
Archives and Research Library	0108 Old Jail Museum and Genealogy Research Center	0.5286
Mary Willis Library	0224 Mary Willis Library	1.0000
Beach Institute		0.0000
Bellevue	0080 LaGrange Woman's Club Bellevue	0.8414
Berry College	0363 Perry Public Library	0.5247
Bethesda Museum		0.0000
Beulah Heights University		0.0000
Blue and Gray Museum	0014 Blue and Gray Museum	1.0000
Bonaventure Cemetery		0.0000
Booth Western Art Museum	0016 Booth Western Art Museum	1.0000
Bowdon Area Historical Society	0017 Bowdon Area Historical Society	1.0000
Brantley Historical Society	0517 Brantley County Library	0.6133
Brasstown Bald Visitor Information Center	0460 Braselton Library	0.6204
Brooks County Genealogical Society	0227 Brooks County Public Library	0.6618
Brooks County Public Library	0227 Brooks County Public Library	1.0000
Bryan-Lang Historical Archives		0.0000
Bulloch County Historical Society	0020 Bulloch County Historical Society	1.0000
Bulloch Hall	0021 Bulloch Hall	1.0000
Butts County Historical Society	0022 Butts County Historical Society	1.0000
Byron Area Historical Society	0017 Bowdon Area Historical Society	0.6133
C.E.Blevins Avian Learning Center		0.0000
Callaway Plantation	0130 Smith Plantation	0.5070
Camden County Historical Commission	0518 Camden County Public Library	0.6660
Candler County Historical Society Museum	0111 Polk County Historical Society Museum	0.5909
Cannonball House	0024 Cannonball House	1.0000
Carroll County Genealogical Society	0037 Decatur County Historical and Genealogical Society	0.6315

Project Update: Documentation

- Instructions and guidelines have been migrated into a single Workflow Manual
- It is growing and changing as our workflow changes (e.g., adding Coordinate Location QC steps)
- Documentation is a gift you give to your future self (esp. if you have a bad memory!)

Workflow Manual

This page describes how to walk through the full GaNCH workflow, so that you can replicate the process for your own region. Links from the GaNCH project are included to provide context, but of course your region may have unique local needs.

Setup

- Download and install free open source software (and one proprietary software) to manipulate data and upload it to Wikidata
 - [Git](#) - [Version Control](#) software, used to protect your work against accidental deletion.
 - [Visual Studio Code](#) - Text and code editor with a built-in terminal. Free and easy to learn interface.
 - [Edit CSV extension](#) - Edit CSV files in spreadsheet format within VS Code. We don't use MS Excel since it reformats content (dates, etc.).
 - [markdownlint extension](#) - Helps with formatting markdown files. We're writing all our documentation in markdown because it provides rich text formatting.
 - [python extension](#) - Python for VS Code.
 - [OpenRefine](#) - Data wrangling software which is also used to upload to Wikidata.
 - [Microsoft Excel](#) - You'll use Excel (or [LibreOffice Calc](#)) to perform data formatting and cleanup with formulas to prepare CSVs. Be careful NOT to use Excel for most editing, as it tends to reformat dates, phone numbers, etc.
 - [Excel Fuzzy Lookup Add-In](#) for de-duplicating the index.

Compile

- Reach out to [partner organizations](#) and individuals in your region to identify [data sources](#).
- Data sources can be found on the web, or may be emailed to you directly from a partner organization. Screen the data for [Personally Identifying Information \(PII\)](#) such as personal email addresses. If the PII is already available publicly on the web (i.e. [stan@organization.org](#) published on the organization's website), we include it. However, if the PII can be replaced with non-personal information (i.e. [info@organization.org](#)), we use the non-personal information.
- Create a [data dictionary](#) that defines what fields you're going to use, how those field match up to the Wikidata schema, and whether they're going to be required, recommended, or optional. This will help you format your data correctly for Wikidata, prioritize work, and explain to other folks what you're doing.
- Reformat the data from each source into its own spreadsheet matching a [CSV template](#).
 - HTML lists from websites (like the GAMAG dataset) are reformatted into CSV by hand using [VS Code's multi-cursor](#) capabilities for batch editing.
 - Straightforward website tables (like the GPLSV dataset) are harvested using the [HTML Table to CSV/Excel Converter](#)

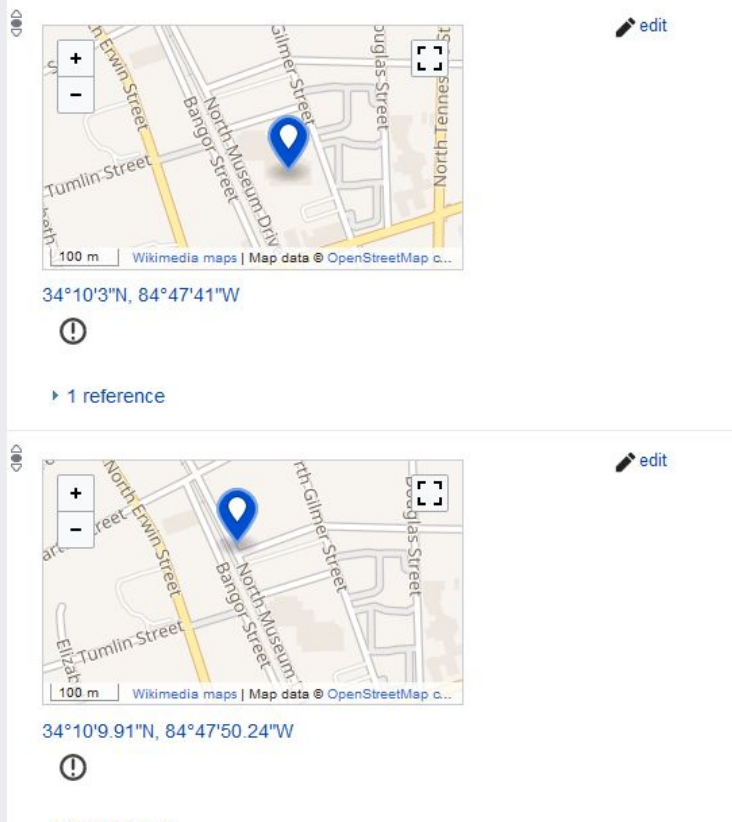
<https://github.com/clifflandis/GaNCH/blob/master/docs/workflow.md>

Challenges: Quality Control

- We're using free services to generate LAT/LONG, and some results are close but not perfect.
- We're performing QC checks on all items post-ingest.
- As we find errors, we're manually correcting them using Google Maps.

<https://www.wikidata.org/wiki/Q4943895>

coordinate location



edit

100 m Wikimedia maps | Map data © OpenStreetMap c...

34°10'3"N, 84°47'41"W

1 reference

edit

100 m Wikimedia maps | Map data © OpenStreetMap c...

34°10'9.91"N, 84°47'50.24"W

2 references

Warning: Small can of worms ahead!

Challenge: P131 should use “most local”

- P131: Located in the Administrative Territorial Entity
- We have been adding city, county, and state (but it should apparently only use the “most local”!)
- These properties are supposed to be transitive and inherit from the most local up to the most global.
- Any mistakes that we made can be undone relatively easily using QuickStatements

Warm Springs (Q995678)

Warm Springs | ウォームスプリング (ジョージア州) | Bullochville
city in Meriwether County, Georgia, United States

Location properties

instance of	City of the United States
	Municipality of Georgia
located in time zone	Eastern Time Zone

Maps



Other Wikidata items within 15km | Geohack | Taginfo | Overpass | 32.8886 / -84.68

Location	
Name	Description
North America	continent on the Earth's northwestern quadrant
United States of America	federal republic in North America
Georgia	state of the United States of America
Meriwether County	county in the state of Georgia, United States
Warm Springs	city in Meriwether County, Georgia, United States

External sites

official website

External sources

archINFORM location	40597
Encyclopædia Britannica	place/Warm-Springs
Freebase	/m/0ry4t
GeoNames	4229468
GNIS	324851
Library of Congress authority	n88109714
MusicBrainz area	b642c70f-e417-42ac-b53b-20e9fcd4db72
New Georgia Encyclopedia	counties-cities-neighborhoods/warm-springs
US National Archives	10040330
VIAF	123289058
VIAF	131453964
Who's on First	85936205

External identifiers

<https://tools.wmflabs.org/reasonator/?q=Q995678>

Challenges: Municipalities

- But some municipalities exist in multiple counties (like Atlanta being in both DeKalb and Fulton counties).
- How do you make sure that an organization shows up in the *correct county* during a query?
- Of the 539 municipalities in Georgia, 51 (9.5 %) are in *more than one county*
- The conversation continues at the link below...

Best practice for representing cities/towns in multiple counties? [\[edit | Add topic \]](#)

Hi folks! I'm trying to figure out the best practice for representing cities and towns that exist in multiple counties. Pull counties: [Allentown \(Q2670381\)](#) is in Bleckley, Laurens, Twiggs, and Wilkinson counties, and [Braselton \(Q899020\)](#) is located in the administrative territorial entity (P131) property, but I didn't know if that's best practice (or if there's con profit organizations in Atlanta, and "the most local admin territory" is the city of Atlanta, would it even be possible to [Clifflandis \(talk\)](#) 18:02, 14 January 2020 (UTC)

[@Clifflandis](#): This is a interesting question. Maybe the solution is to create a item for the part of Atlanta in Fultor

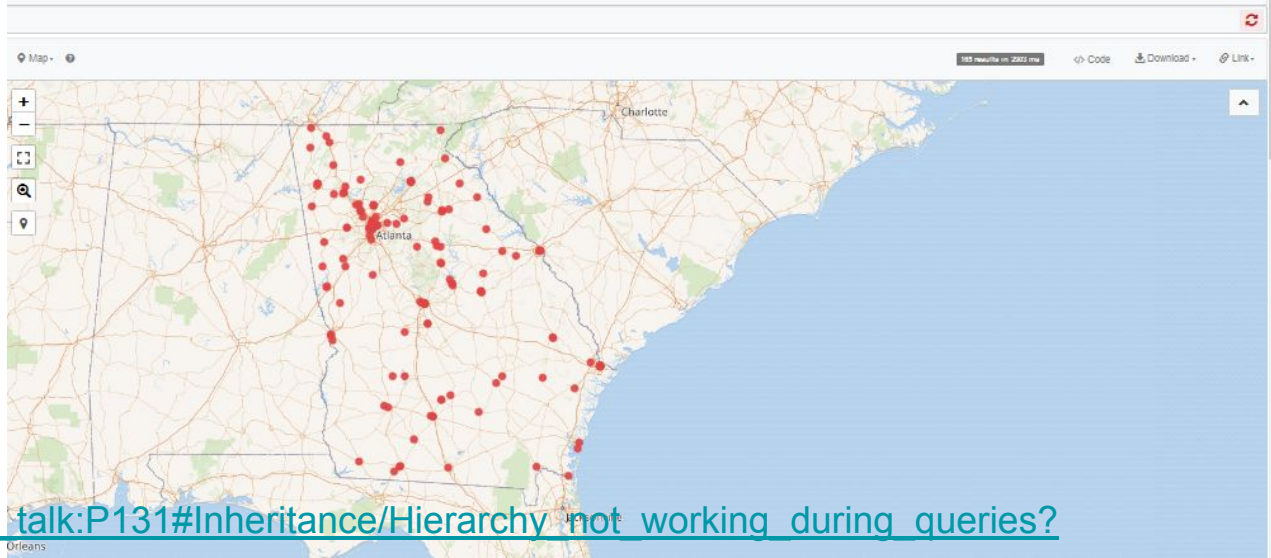
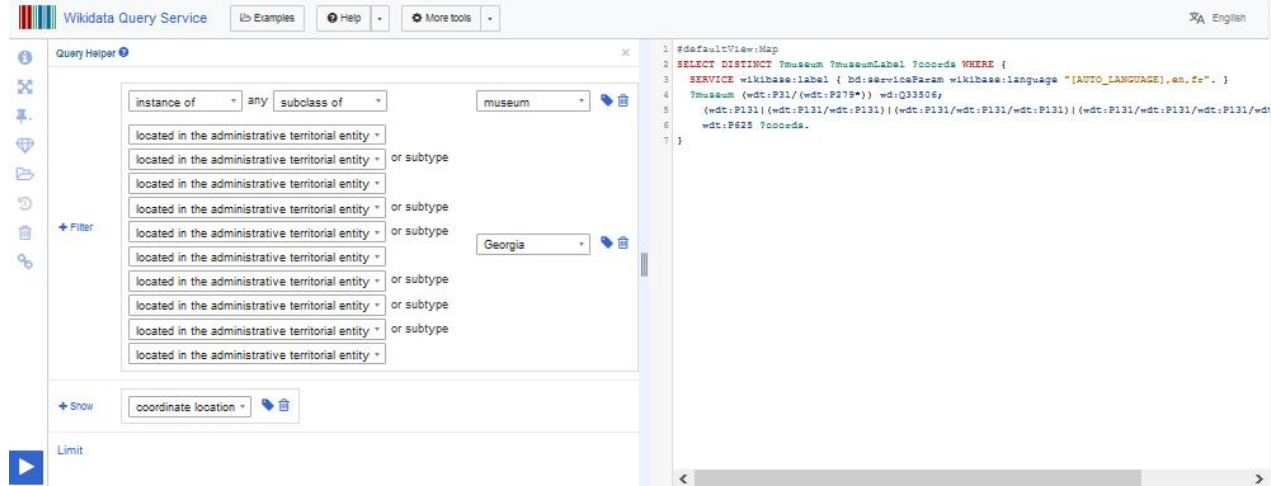


https://www.wikidata.org/wiki/Property_talk:P131#Best_practice_for_representing_cities/towns_in_multiple_counties?

<https://tools.wmflabs.org/reasonator/?&q=2670381>

Challenge: Constructing Queries

- With guidance from the Wikidata community, we'll be constructing queries that make use of hierarchical, transitive inheritance ...once we get this municipalities thing sorted out!



https://www.wikidata.org/wiki/Property_talk:P131#Inheritance/Hierarchy_not_working_during_queries?

Challenges? No problem!

- We're learning from the Wikidata community, and they're learning from us (as we find problems)!
- Wikidata will be celebrating its [8th birthday](#) this Fall, so these are learning/growing moments.
- Regardless of how we need to structure the queries, we'll make sure that the website is pulling the correct data.



https://www.androidcentral.com/sites/androidcentral.com/files/styles/large/public/article_images/2015/08/dont-panic.jpg?itok=Eskz_M2z

Museum map results @ last meeting: 82 results

Wikidata Query Service

Examples Help More tools

English

Query Helper

Filter

- instance of any
- subclass of
- located in the administrative territorial entity Georgia

Show

coordinate location

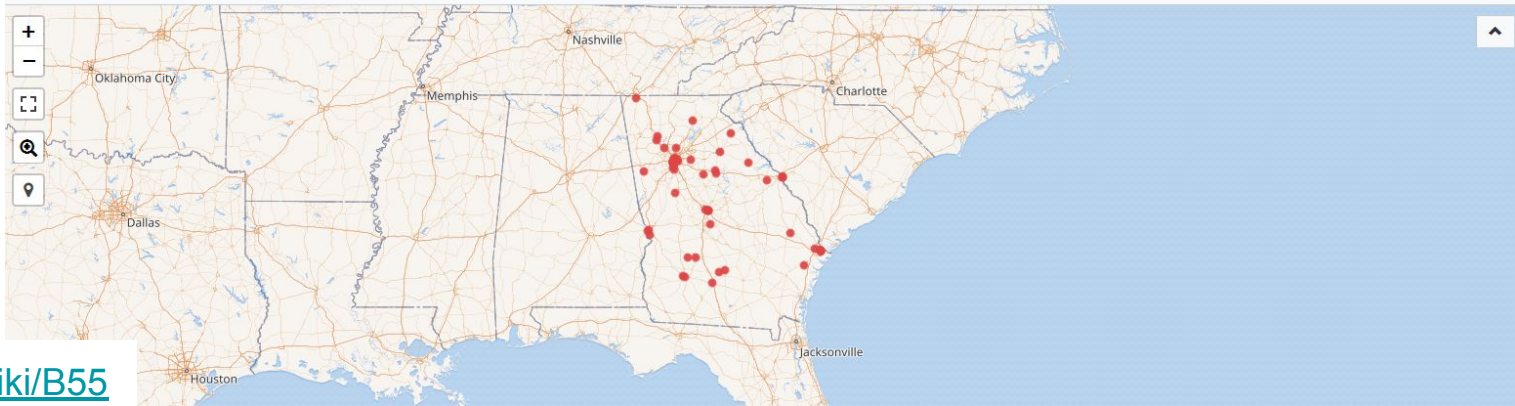
Limit 10000

```
1 #Map of Libraries in Canada
2 # Canadian libraries on a map (must have coordinates!)
3 #defaultView:Map
4 SELECT ?library ?libraryLabel ?coords WHERE {
5   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en,fr". }
6   ?library (wdt:F31/(wdt:P279*)) wd:Q33506;
7   wdt:F131 wd:Q1428;
8   wdt:P625 ?coords.
9 }
10 LIMIT 10000
```

Map

82 results in 3985 ms

Code Download Link



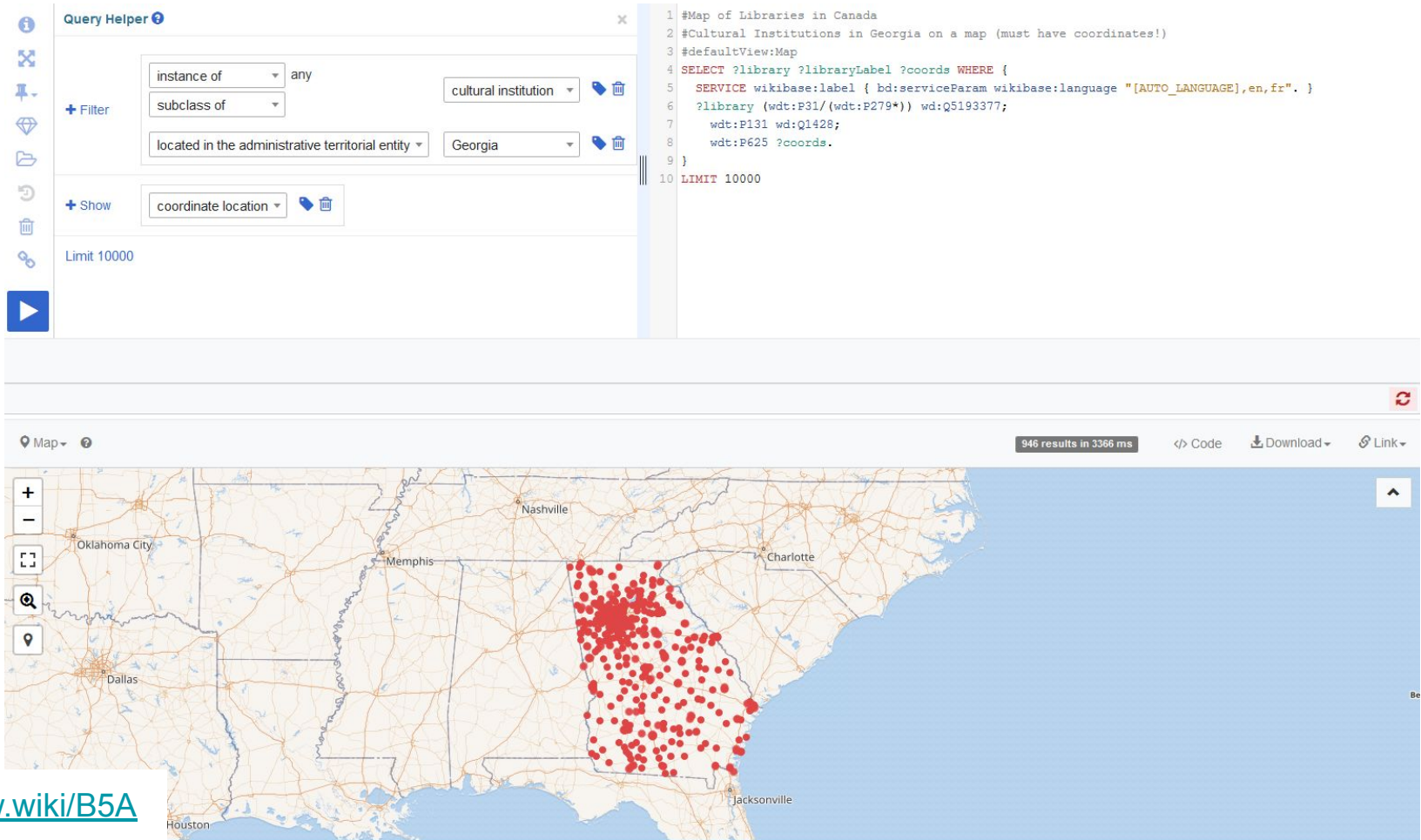
Museum map results now: 165 results!

The screenshot displays the Wikidata Query Service interface. The top navigation bar includes links for Wikidata Query Service, Examples, Help, and More tools. The main interface is divided into two panels. The left panel, titled 'Query Helper', contains a query editor with a dropdown menu for 'instance of' set to 'any', a dropdown for 'subclass of' set to 'museum', and a dropdown for 'located in the administrative territorial entity' set to 'Georgia'. Below these are several 'or subtype' options. The right panel shows the SPARQL query:

```
#defaultView:Map
SELECT DISTINCT ?museum ?museumLabel ?coords WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en,en-gb" . }
  ?museum (wdt:P31/(wdt:P279*)) wdt:Q33506;
  (wdt:P31)/(wdt:P31/(wdt:P31/(wdt:P31))) (wdt:P31/(wdt:P31/(wdt:P31)))
  wdt:P625 ?coords.
}
```

 The bottom panel shows a map of Georgia with numerous red dots representing museum locations. The map includes labels for Atlanta, Charlotte, and Jacksonville. The bottom status bar indicates '10 results in 200 ms' and provides links for Code, Download, and Link.

Cultural Institutions in GA with coordinates @ last meeting: 946



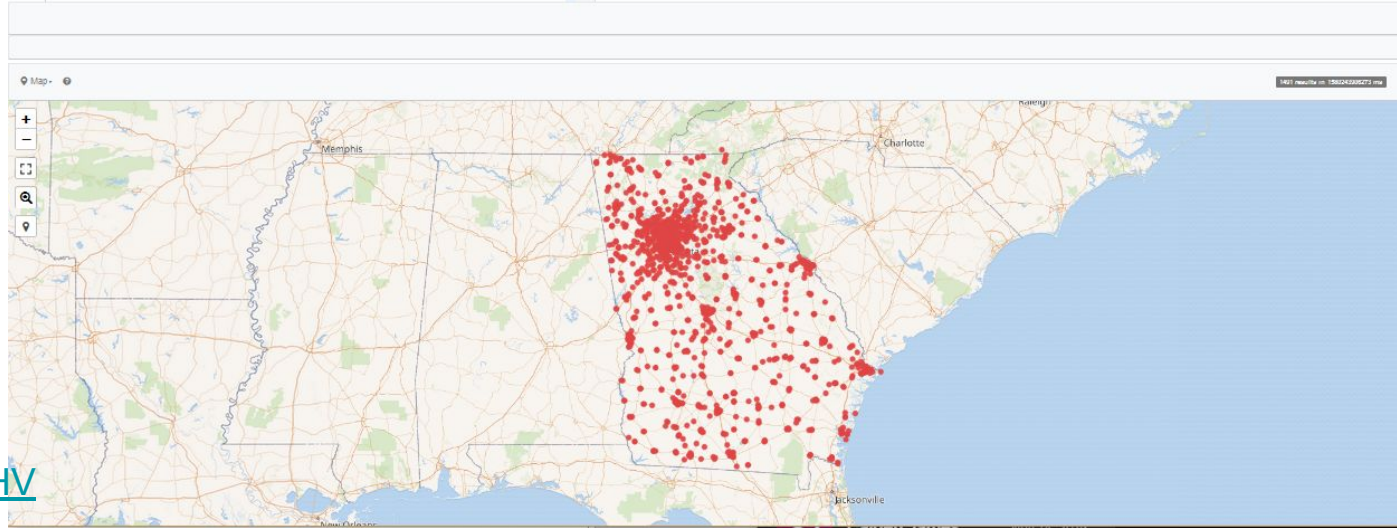
Cultural Institutions in GA with coordinates now: 1,491!

The screenshot displays the Wikidata Query Service interface. On the left, the "Query Helper" sidebar shows a complex filter chain: "instance of" -> "any" -> "subclass of" -> "cultural institution". Below this, there are multiple instances of "located in the administrative territorial entity" connected by "or subtype" relationships. A "Filter" button is visible. At the bottom, a "Show" button is next to a "coordinate location" filter.

The main area contains a SPARQL query:

```
1 #defaultView:Map
2 SELECT DISTINCT ?museum ?museumLabel ?coords WHERE {
3   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en,fr". }
4   ?museum (wdt:P31/(wdt:P279*)) wd:Q5193377;
5           (wdt:P131|(wdt:P131/wdt:P131)|(wdt:P131/wdt:P131/wdt:P131)|(wdt:P131/wdt:P131/wdt:P131/wdt:P131) wd:Q1428;
6           wdt:P625 ?coords.
7 }
```

On the right, the first result is shown as a map view, displaying a museum building and its location on a map of Georgia.



Up Next...

- Continue to prep, ingest, and QC datasets
- Continue to dialogue with Wikidata community to address challenges
- Begin planning forthcoming website for emergency responders' access
- Next Meeting:

Thursday, March 26, 2020

Shortcode	Organization	Source Name	Link	# Items
GHRAC	Georgia Historical Records Advisory Council	Directory of Historical and Cultural Organizations	https://apps.georgiaarchives.org/GHRAB/	641
GHSOC	Georgia Historical Society	Affiliate Chapter List by County	https://georgiahistory.com/education-outreach/affiliate-chapter-program/affiliate-chapter-list-by-county/	44+36+47 = 127
GAMAG	Georgia Association of Museums and Galleries	Museum Links	http://www.gamg.org/museumlinks.html	155
GPLSV	Georgia Public Library Service	All Public Library Facilities	https://georgialibraries.org/allpubliclibraryfacilities/	392
GALIL	GALILEO	Participating Institutions	https://about.galileo.usg.edu/institutions	93
TCSOG	Technical College System of Georgia	Find a College	https://tcsog.edu/find-a-college/	22
ARTGA	ArtsGeorgia	Directory of Directories	https://www.artsgeorgia.net/directory/directory-of-directoriescalendars/	Meta
GSCCC	Georgia Superior Court Clerks' Cooperative Authority	Clerks	https://www.gscca.org/Clerks/	159
GINFO	GeorgialInfo	Counties	https://georgiainfo.galileo.usg.edu/topics/counties/	159
IMLSS	Institute of Museum and Library Services	Georgia Museums List	https://docs.google.com/spreadsheets/d/104wXi8wh6VXlJfJNYlRtph9G7J7TxztVD-ZjhGF70/	491
GADNR	Georgia Department of Natural Resources	Georgia State Parks, Historic Sites, and Partners	https://gastateparks.org/Map	72
GCFTA	Georgia Council for the Arts	FY2020 Grant Applicants	N/A (emailed)	556

Questions?

